Inverse Problems and Imaging doi:10.3934/ipi.2020067
Volume 15, No. 1, 2021, 109–128

FAST ALGORITHMS FOR ROBUST PRINCIPAL COMPONENT ANALYSIS WITH AN UPPER BOUND ON THE RANK

Ningyu Sha

Department of Computational Mathematics, Science and Engineering
Michigan State University, East Lansing, MI 48824, USA

Lei Shi

School of Mathematical Sciences
Shanghai Key Laboratory for Contemporary Applied Mathematics
Key Laboratory of Mathematics for Nonlinear Sciences (Fudan University)
Ministry of Education
Fudan University, Shanghai, 200433, China

Ming Yan*

Department of Computational Mathematics
Science and Engineering, Department of Mathematics
Michigan State University, East Lansing, MI 48824, USA

Abstract. The robust principal component analysis (RPCA) decomposes a data matrix into a low-rank part and a sparse part. There are mainly two types of algorithms for RPCA. The first type of algorithm applies regularization terms on the singular values of a matrix to obtain a low-rank matrix. However, calculating singular values can be very expensive for large matrices. The second type of algorithm replaces the low-rank matrix as the multiplication of two small matrices. They are faster than the first type because no singular value decomposition (SVD) is required. However, the rank of the low-rank matrix is required, and an accurate rank estimation is needed to obtain a reasonable solution. In this paper, we propose algorithms that combine both types. Our proposed algorithms require an upper bound of the rank and SVD on small matrices. First, they are faster than the first type because the cost of SVD on small matrices is negligible. Second, they are more robust than the second type because an upper bound of the rank instead of the exact rank is required. Furthermore, we apply the Gauss-Newton method to increase the speed of our algorithms. Numerical experiments show the better performance of our proposed algorithms.

1. Introduction.

Robust principal component analysis (RPCA) decomposes a data matrix into a low-rank part and a sparse part. It has applications in a wide range of areas, including computer vision [8], image processing [16, 9], dimensionality reduction [6], and bioinformatics data analysis [7]. More specifically, the RPCA model has achieved great success in video surveillance and face recognition [4, 2].

2020 *Mathematics Subject Classification.* Primary: 65K10, 90C26; Secondary: 65D18.

Key words and phrases. Robust principal component analysis, nonconvex, acceleration, low-rank, sparse.

N. Sha and M. Yan are supported by NSF grant DMS-1621798 and DMS-2012439. L. Shi is supported by NNSFC grant 11631015 and Shanghai Science and Technology Research Program 19JC1420101 and and 20JC1412700.

* Corresponding author.

For example, in video surveillance, the low-rank part preserves the stationary background, whereas the sparse part can capture a moving object or person in the foreground.

We first assume that the data matrix $\mathbf{D}$ is obtained by the sum of a low-rank matrix and a sparse matrix. That is

$$\mathbf{D} = \mathbf{L} + \mathbf{S},$$

where $\mathbf{L}$ is a low-rank matrix and $\mathbf{S}$ is a sparse matrix having only a few nonzero entries. RPCA is an inverse problem to recover $\mathbf{L}$ and $\mathbf{S}$ from the matrix $\mathbf{D}$, which can be realized via solving the idealized nonconvex problem

$$\underset{\mathbf{L},\mathbf{S}}{\text{minimize}}\ \ \text{rank}(\mathbf{L}) + \lambda\|\mathbf{S}\|_0,\ \text{subject to } \mathbf{L}+\mathbf{S}=\mathbf{D}, \tag{1}$$

where λ is a parameter to balance the two objectives and $\|\mathbf{S}\|_0$ counts the number of non-zero entries in $\mathbf{S}$. However, this problem is NP-hard in general [1]. Therefore, much attention is focused on the following convex relaxation:

$$\underset{\mathbf{L},\mathbf{S}}{\text{minimize}}\ \ \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1,\ \text{subject to } \mathbf{L}+\mathbf{S}=\mathbf{D}. \tag{2}$$

Here $\|\cdot\|_*$ and $\|\cdot\|_1$ denote the nuclear norm and ℓ_1-norm of a matrix, respectively. It is shown that under mild conditions, the convex model (2) can exactly recover the low-rank and sparse parts with high probabilities [4]. When additional Gaussian noise is considered, we can set the noise level to be ϵ and use the Frobenius norm $\|\cdot\|_F$ to measure the reconstruction error. Then, the problem becomes

$$\underset{\mathbf{L},\mathbf{S}}{\text{minimize}}\ \ \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1,\ \text{subject to } \|\mathbf{L}+\mathbf{S}-\mathbf{D}\|_F^2 \leq \epsilon. \tag{3}$$

This constrained optimization problem is equivalent to the unconstrained problem

$$\underset{\mathbf{L},\mathbf{S}}{\text{minimize}}\ \ \frac{1}{2}\|\mathbf{L}+\mathbf{S}-\mathbf{D}\|_F^2 + \mu\|\mathbf{L}\|_* + \lambda\mu\|\mathbf{S}\|_1 \tag{4}$$

with a trade-off parameter μ. There is a correspondence between the two parameters ϵ and μ in (3) and (4), but the explicit expression does not exist. In this paper, we will focus on the unconstrained problem (4), and the technique introduced in this paper can be applied to the convex models (2) and (3). Please see Section 4 for more details.

There are many existing approaches for solving (4), including the augmented Lagrange method [15, 2, 25]. Some examples are proximal gradient method for $(\mathbf{L},\mathbf{S})$, alternating minimization for $\mathbf{L}$ and $\mathbf{S}$ [20], proximal gradient method for $\mathbf{L}$ after $\mathbf{S}$ is eliminated [19], alternating direction method of multipliers (ADMM) [26, 21]. All these approaches need to find the proximal of the nuclear norm, which requires singular value decomposition (SVD). When the matrix size is large, the SVD computation is very expensive and dominates other computation [22].

Alternative approaches for RPCA use matrix decomposition [24] and do not require SVD. Assuming that the rank of $\mathbf{L}$ is known as p, we can decompose it as

$$\mathbf{L} = \mathbf{X}\mathbf{Y}^\top,$$

with $\mathbf{X} \in \mathbb{R}^{m\times p}$ and $\mathbf{Y} \in \mathbb{R}^{n\times p}$. Then the following nonconvex optimization problem

$$\underset{\mathbf{X},\mathbf{Y},\mathbf{S}}{\text{minimize}}\ \ \frac{1}{2}\|\mathbf{X}\mathbf{Y}^\top + \mathbf{S}-\mathbf{D}\|_F^2 + \lambda\|\mathbf{S}\|_1, \tag{5}$$

is considered. There are infinite many optimal solutions for this problem, since for any invertable matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, $(\mathbf{X},\mathbf{Y},\mathbf{S})$ and $(\mathbf{X}\mathbf{A}^{-1},\mathbf{Y}\mathbf{A}^\top,\mathbf{S})$ have the same objective value. In fact, for any matrix $\mathbf{L}$ with rank no greater than p, we can find $\mathbf{L} = \mathbf{X}\mathbf{Y}^\top$ with $\mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}$. Therefore, we can have an additional constraint $\mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}$. The resulting problem still has infinite many optimal solutions, since for any orthogonal matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, $(\mathbf{X}, \mathbf{Y}, \mathbf{S})$ and $(\mathbf{XA}, \mathbf{YA}, \mathbf{S})$ have the same objective value. Though $(\mathbf{X}, \mathbf{Y})$ are not unique, the low-rank matrix $\mathbf{L} = \mathbf{X}\mathbf{Y}^\top$ that we need could be unique. This resulting problem was discussed in [20], and an efficient algorithm by alternating minimizing $\mathbf{X}\mathbf{Y}^\top$ and $\mathbf{S}$ is provided. In this algorithm, a Gauss-Newton algorithm is applied to update $\mathbf{X}\mathbf{Y}^\top$ and reduce the time.

Though the matrix decomposition approach could be solved faster than the nuclear norm minimization approach because no SVD is required, it is nonconvex and requires an accurate estimation of the rank of $\mathbf{L}$. Fig. 2 in Section 3.1.2 demonstrates that a good estimation of the rank is critical. However, in most scenarios, we do not have the exact rank of $\mathbf{L}$, but we can have an upper bound of the true rank. Therefore, we can combine the matrix decomposition and the nuclear norm minimization to have the benefits of both approaches: fast speed and robustness in the rank. The problem we consider in this paper is

$$\text{(6)} \qquad \underset{\mathbf{L},\mathbf{S}}{\text{minimize}} \ \frac{1}{2}\|\mathbf{L}+\mathbf{S}-\mathbf{D}\|_F^2 + \mu\|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1, \ \text{subject to } \operatorname{rank}(\mathbf{L}) \le p.$$

When $\mu = 0$, the problem (6) is equivalent to (5). In addition, we consider the following more general problem

$$\text{(7)} \qquad \underset{\mathbf{L},\mathbf{S}}{\text{minimize}} \ \frac{1}{2}\|\mathcal{A}(\mathbf{L})+\mathbf{S}-\mathbf{D}\|_F^2 + \mu\|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1, \ \text{subject to } \operatorname{rank}(\mathbf{L}) \le p,$$

where $\mathbf{D}$ is the measurement of $\mathcal{A}(\mathbf{L})$ contaminated with both Gaussian noise and sparse noise. Here $\mathcal{A}$ is a bounded linear operator that describes how the measurements are calculated. For example, in robust matrix completion, we let $\mathcal{A}$ be the restriction operator on the given components of the matrix $\mathbf{L}$.

Note that the alternating minimization algorithm in [20] can not be applied to this general problem because the subproblem for $\mathbf{L}$ can no longer be solved efficiently by the Gauss-Newton method. We will show the equivalency of the alternating minimization algorithm in [20] and a proximal gradient method applied to a problem with $\mathbf{L}$ only. Then the subproblem of $\mathbf{L}$ in our general problem (7) can still be solved efficiently with the Gauss-Newton method. Please see more details in Section 2.

For simplicity, we use the nuclear norm and ℓ_1-norm for the low-rank and sparse matrices, respectively. The main purpose of this paper is to introduce a fast algorithm to solve (7). Though the technique can be applied to variants of (7), as will be shown in Section 4, the comparison of different penalties is out of the scope of this paper. The contributions of this paper are:

- We propose a new model for RPCA, which combines the nuclear norm minimization and the matrix decomposition. The matrix decomposition brings efficient algorithms, and the nuclear norm minimization on a smaller matrix removes the requirement of the rank of the low-rank matrix. Note that other nonconvex penalties can replace the nuclear norm minimization, and the results in this paper are still valid.
- We develop efficient algorithms using Gauss-Newton to solve this problem and show its convergence.

1.1. **Notation.** Throughout this paper, matrices are denoted by bold capital letters (e.g., $\mathbf{A}$), and operators are denoted by calligraphic letters (e.g., $\mathcal{A}$). In particular, $\mathbf{I}$ denotes the identity matrix, $\mathbf{0}$ denotes the zero matrix (all entries equal zero), and $\mathcal{I}$ denotes the identity operator. If there is potential for confusion, we indicate the dimension of matrix with subscripts. For a matrix $\mathbf{A}$, $\mathbf{A}^\top$ represents its transpose and $\mathbf{A}(:, j : k)$ denotes the matrix composed by the columns of $\mathbf{A}$ indexing from j to k. Let $\mathbf{A}_{i,j}$ be the (i, j) entry of $\mathbf{A}$. The ℓ_1-norm of $\mathbf{A}$ is given by $\|\mathbf{A}\|_1 = \sum_{i,j} |\mathbf{A}_{i,j}|$. We denote the ith singular value of $\mathbf{A}$ by $\sigma_i(\mathbf{A})$. The nuclear norm of $\mathbf{A}$ is given by $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$. We will use $\partial\|\cdot\|_1$ and $\partial\|\cdot\|_*$ to denote the subgradients of ℓ_1-norm and nuclear norm, respectively. The linear space of all $m \times n$ real matrices is denoted by $\mathbb{R}^{m\times n}$. For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, the inner product of $\mathbf{A}, \mathbf{B}$ is defined by $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{Tr}(\mathbf{A}^\top\mathbf{B})$, which induces the Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\mathrm{Tr}(\mathbf{A}^\top\mathbf{A})} = \sqrt{\sum_i \sigma_i^2(\mathbf{A})}$. Let $\mathcal{A}$ be a linear bounded operator on $\mathbb{R}^{m\times n}$. The operator norm of $\mathcal{A}$ is given by $\|\mathcal{A}\| = \sup\{\|\mathcal{A}(\mathbf{A})\|_F : \mathbf{A} \in \mathbb{R}^{m\times n}, \|\mathbf{A}\|_F = 1\}$. The adjoint operator of $\mathcal{A}$ denoted by $\mathcal{A}^*$ is also linear and bounded on $\mathbb{R}^{m\times n}$ such that $\langle\mathcal{A}(\mathbf{A}), \mathbf{B}\rangle = \langle\mathbf{A}, \mathcal{A}^*(\mathbf{B})\rangle$. Notation $\odot$ is used to denote the component-wise multiplication. Additionally, for a function $f : \mathbb{R} \to \mathbb{R}$, without further reference, f acting on a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ specifies that f is evaluated on each entry of $\mathbf{A}$, i.e., $f(\mathbf{A}) \in \mathbb{R}^{m\times n}$ with $(f(\mathbf{A}))_{i,j} = f(\mathbf{A}_{i,j})$. For example, if $f(x) = |x| - \lambda$, we can denote $f(\mathbf{A}) \in \mathbb{R}^{m\times n}$ by $|\mathbf{A}| - \lambda$ with $(|\mathbf{A}| - \lambda)_{i,j} = |\mathbf{A}_{i,j}| - \lambda$.

1.2. **Organization.** The rest of the paper is organized as follows. We introduce our proposed algorithms and show their convergence in Section 2. Then we conduct numerical experiments to compare our proposed algorithms' performance with existing approaches in Section 3. In Section 4, we conclude this paper with some potential extensions.

2. Proposed algorithms.

The problem (6) is nonconvex because of the constraint $\mathrm{rank}(\mathbf{L}) \le p$. It has several equivalent formulations. E.g., it is equivalent to the following nonconvex weighted nuclear norm minimization problem:

$$\underset{\mathbf{L},\mathbf{S}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{L}+\mathbf{S}-\mathbf{D}\|_F^2 + \mu\sum_{i=1}^{p}\sigma_i(\mathbf{L}) + C\sum_{i=p+1}^{\min(m,n)}\sigma_i(\mathbf{L}) + \lambda\|\mathbf{S}\|_1,$$

where C is a sufficiently large number such that the optimal $\mathbf{L}$ has at most p nonzero singular values. However, this formulation also requires the singular value decomposition of a $m \times n$ matrix in each iteration, which is expensive when m and n are large. We consider another equivalent problem with matrix decomposition in the following theorem.

Theorem 2.1. *Problem (6) is equivalent to*

$$\underset{\mathbf{X},\mathbf{Y},\mathbf{S}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{X}\mathbf{Y}^\top+\mathbf{S}-\mathbf{D}\|_F^2 + \mu\|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_1,\ \text{subject to } \mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}. \tag{8}$$

More specifically, if $(\mathbf{X}, \mathbf{Y}, \mathbf{S})$ is an optimal solution to (8), then $(\mathbf{X}\mathbf{Y}^\top, \mathbf{S})$ is an optimal solution to (6). If $(\mathbf{L}, \mathbf{S})$ is an optimal solution to (6) and we have the decomposition $\mathbf{L} = \mathbf{X}\mathbf{Y}^\top$ with $\mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}$, then $(\mathbf{X}, \mathbf{Y}, \mathbf{S})$ is an optimal solution to (8).

Proof. For any matrix $\mathbf{L} \in \mathbb{R}^{m\times n}$ with rank no greater than p, we can have the decomposition

$$\mathbf{L} = \mathbf{X}\mathbf{Y}^\top,$$

with $\mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}$. This decomposition is not unique, and one decomposition can be easily obtained from the compact SVD of $\mathbf{L}$. Let $\mathbf{L} = \mathbf{U}_p\Sigma_p\mathbf{V}_p^\top$ be the SVD of $\mathbf{L}$ with a square $p \times p$ matrix Σ_p, we have $\mathbf{V}_p^\top\mathbf{V}_p = \mathbf{I}_{p\times p}$. Thus, problem (6) is equivalent to

$$\underset{\mathbf{X},\mathbf{Y},\mathbf{S}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{X}\mathbf{Y}^\top + \mathbf{S} - \mathbf{D}\|_F^2 + \mu\|\mathbf{X}\mathbf{Y}^\top\|_* + \lambda\|\mathbf{S}\|_1,\ \text{subject to } \mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}.$$

For any $\mathbf{X} \in \mathbb{R}^{m\times p}$, let $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top$ be its SVD with $\mathbf{U} \in \mathbb{R}^{m\times p}$ and $\mathbf{V} \in \mathbb{R}^{p\times p}$. We have

$$\mathbf{X}\mathbf{Y}^\top = \mathbf{U}\Sigma\mathbf{V}^\top\mathbf{Y}^\top = \mathbf{U}\Sigma(\mathbf{Y}\mathbf{V})^\top.$$

Since $(\mathbf{Y}\mathbf{V})^\top(\mathbf{Y}\mathbf{V}) = \mathbf{V}^\top\mathbf{Y}^\top\mathbf{Y}\mathbf{V} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}_{p\times p}$. The SVD of $\mathbf{X}\mathbf{Y}^\top$ is $\mathbf{U}\Sigma(\mathbf{Y}\mathbf{V})^\top$, and $\|\mathbf{X}\mathbf{Y}^\top\|_* = \sum_{i=1}^p \Sigma_{ii} = \|\mathbf{X}\|_*$. Thus, problem (6) is equivalent to (8). □

Next, we consider problem (8) with $\mathbf{S}$ fixed. When $\mathbf{S}$ is fixed, it becomes a problem of $\mathbf{L} = \mathbf{X}\mathbf{Y}^\top$, and solving this problem is to find the proximal operator of the corresponding nonconvex weighted nuclear norm, which is denoted as

$$\underset{\mathbf{L}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{L} - \mathbf{M}\|_F^2 + \mu\|\mathbf{L}\|_*,\ \text{subject to } \operatorname{rank}(\mathbf{L}) \le p, \tag{9}$$

or equivalently

$$\underset{\mathbf{X},\mathbf{Y}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{X}\mathbf{Y}^\top - \mathbf{M}\|_F^2 + \mu\|\mathbf{X}\|_*,\ \text{subject to } \mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}, \tag{10}$$

where $\mathbf{M} = \mathbf{D} - \mathbf{S}$.

Theorem 2.2. *Let $q = \min(m, n)$. Problem (9) can be solved in two steps:*

1. *Find the compact SVD of $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\top$, with $\Sigma = \operatorname{diag}(\sigma_1(\mathbf{M}), \cdots, \sigma_q(\mathbf{M}))$ satisfying $\sigma_1(\mathbf{M}) \ge \sigma_2(\mathbf{M}) \ge \cdots \ge \sigma_q(\mathbf{M})$;*
2. *Construct a diagonal matrix $\hat{\Sigma}_\mu \in \mathbb{R}^{p\times p}$ with $(\hat{\Sigma}_\mu)_{ii} = \max(\Sigma_{ii} - \mu, 0)$, then one solution of (9) is $\mathbf{U}(:, 1:p)\hat{\Sigma}_\mu\mathbf{V}(:, 1:p)^\top$.*

In addition, for any orthogonal matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, $(\mathbf{U}(:, 1:p)\hat{\Sigma}_\mu\mathbf{A}, \mathbf{V}(:, 1:p)\mathbf{A})$ is an optimal solution of (10).

Proof. Given any $\mathbf{L} \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(\mathbf{L}) \le p$, let $\sigma_1, \sigma_2, \cdots, \sigma_q$ be its singular values in the decreasing order such that $\sigma_{p+1} = \cdots = \sigma_q = 0$. Note that the main diagonal entries of Σ are the singular values of $\mathbf{M}$. According to von-Neumann trace inequality [11, Theorem 7.4.1.1], one can bound the matrix inner product by the singular values, i.e., $\langle\mathbf{L}, \mathbf{M}\rangle \le \sum_{i=1}^q \sigma_i\Sigma_{ii}$. Then we have

$$\begin{aligned}
\frac{1}{2}\|\mathbf{L} - \mathbf{M}\|_F^2 + \mu\|\mathbf{L}\|_* =& \frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{M}\|_F^2 - \langle\mathbf{L}, \mathbf{M}\rangle + \mu\|\mathbf{L}\|_* \\
\ge& \frac{1}{2}\sum_{i=1}^q \sigma_i^2 + \frac{1}{2}\sum_{i=1}^q \Sigma_{ii}^2 - \sum_{i=1}^q \sigma_i\Sigma_{ii} + \mu\sum_{i=1}^q \sigma_i \\
=& \frac{1}{2}\sum_{i=1}^p \sigma_i^2 + \frac{1}{2}\sum_{i=1}^q \Sigma_{ii}^2 - \sum_{i=1}^p \sigma_i\Sigma_{ii} + \mu\sum_{i=1}^p \sigma_i,
\end{aligned} \tag{11}$$

where the equality is satisfied when $\mathbf{L}$ has a simultaneous SVD with $\mathbf{M}$ through $\mathbf{U}$ and $\mathbf{V}$. Therefore, the optimal $\mathbf{L}$ minimizing $\frac{1}{2}\|\mathbf{L} - \mathbf{M}\|_F^2 + \mu\|\mathbf{L}\|_*$ can be selected from the matrices that have a simultaneous SVD with $\mathbf{M}$ through $\mathbf{U}$ and $\mathbf{V}$. Then we can assume that the optimal $\mathbf{L}$ satisfies

$$\mathbf{L} = \mathbf{U}\operatorname{diag}(\sigma_1, \cdots, \sigma_p, \sigma_{p+1}, \cdots, \sigma_q)\mathbf{V}^\top = \mathbf{U}(:, 1:p)\operatorname{diag}(\sigma_1, \cdots, \sigma_p)\mathbf{V}(:, 1:p)^\top,$$

where the last equality holds because of the fact that $\sigma_{p+1} = \cdots = \sigma_q = 0$. Next, one can construct an optimal $\mathbf{L}$ of the above form by letting $\sigma_i = \max(\Sigma_{ii} - \mu, 0)$ for $i = 1, 2, \cdots, p$, which minimizes the last equation in (11). Thus $\mathbf{U}(:, 1 : p)\hat{\Sigma}_\mu \mathbf{V}(:, 1 : p)^\top$ minimizes the objective function of (9) over all $\mathbf{L} \in \mathbb{R}^{m\times n}$ with rank no greater than p.

By the same argument in the proof of Theorem 2.1, we see that problem (10) is equivalent to problem (9). Since for any orthogonal matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, there hold

$$\mathbf{L} = (\mathbf{U}(:, 1 : p)\hat{\Sigma}_\mu \mathbf{A})(\mathbf{V}(:, 1 : p)\mathbf{A})^\top$$

and

$$(\mathbf{V}(:, 1 : p)\mathbf{A})^\top(\mathbf{V}(:, 1 : p)\mathbf{A}) = \mathbf{A}^\top\mathbf{A} = \mathbf{I}_{p\times p}.$$

Therefore, $(\mathbf{U}(:, 1 : p)\hat{\Sigma}_\mu \mathbf{A}, \mathbf{V}(:, 1 : p)\mathbf{A})$ is an optimal solution of problem (10). □

The first step to solve problem (10) in the previous theorem requires the truncated SVD of an $m\times n$ matrix $\mathbf{M}$. Since we only need the first p ($p < q = \min(m, n)$) singular values, we use the Gauss-Newton algorithm to find $(\mathbf{X}, \mathbf{Y})$ alternatively. In this approach, we require the SVD of a $m \times p$ matrix, which is much faster than the truncated SVD of a $m \times n$ matrix when p is small. In addition, we use the previous $\mathbf{X}$ as the initial guess in the next iteration to reduce the number of inner iterations for the Gauss-Newton algorithm.

Lemma 2.3. *If the rank of $\mathbf{M} \in \mathbb{R}^{m\times n}$ is larger than p, problem (10) can be solved in the following three steps:*

1. *Find $\hat{\mathbf{X}} \in \mathbb{R}^{m\times p}$ ($p < m$) by solving the following optimization problem*

$$\operatorname*{minimize}_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{X}\mathbf{X}^\top - \mathbf{M}\mathbf{M}^\top\|_F^2;$$

2. $\mathbf{Y} = \mathbf{M}^\top\hat{\mathbf{X}}(\hat{\mathbf{X}}^\top\hat{\mathbf{X}})^{-1}$*;*
3. *Let $\hat{\mathbf{X}} = \mathbf{U}_p\hat{\Sigma}\mathbf{A}$ be its thin SVD with $\hat{\Sigma} \in \mathbb{R}^{p\times p}$ and choose $\mathbf{X}$ as $\mathbf{X} = \mathbf{U}_p\hat{\Sigma}_\lambda\mathbf{A}$ with $(\hat{\Sigma}_\lambda)_{ii} = \max(0, \hat{\Sigma}_{ii} - \mu)$ for $i = 1, \ldots, p$. Then $(\mathbf{X}, \mathbf{Y})$ is an solution of problem (10).*

Proof. Given any $\mathbf{X} \in \mathbb{R}^{m\times p}$, let $\lambda_1, \lambda_2, \cdots, \lambda_m$ be the non-negative eigenvalues of the matrix $\mathbf{X}\mathbf{X}^\top$. Since $\operatorname{rank}(\mathbf{X}) \le p < m$, we have $\lambda_{p+1} = \cdots = \lambda_m = 0$. Recall that the compact SVD of $\mathbf{M}$ given in Theorem 2.2 is $\mathbf{U}\Sigma\mathbf{V}^\top$ with $\Sigma \in \mathbb{R}^{q\times q}$ (here $q = \min(m, n)$). Then $\Sigma_{11}^2 \ge \Sigma_{11}^2 \ge \cdots \ge \Sigma_{qq}^2$ are the largest q eigenvalues of the matrix $\mathbf{M}\mathbf{M}^\top$, and if $q < m$, the remaining eigenvalues of $\mathbf{M}\mathbf{M}^\top$ are all zeros. Then we have

$$\begin{aligned}\|\mathbf{X}\mathbf{X}^\top - \mathbf{M}\mathbf{M}^\top\|_F^2 &\ge \sum_{i=1}^p \lambda_i^2 + \sum_{i=1}^q \Sigma_{ii}^4 - 2\sum_{i=1}^p \lambda_i\Sigma_{ii}^2\\ &= \sum_{i=1}^p(\lambda_i - \Sigma_{ii}^2)^2 + \sum_{i=p+1}^q \Sigma_{ii}^4 \ge \sum_{i=p+1}^q \Sigma_{ii}^4,\end{aligned}$$

where the equality is satisfied when we choose $\mathbf{X} = \mathbf{U}(:, 1 : p)\operatorname{diag}(\Sigma_{11}, \cdots, \Sigma_{pp})$. Let $\hat{\Sigma} = \operatorname{diag}(\Sigma_{11}, \cdots, \Sigma_{pp})$. The matrix $\hat{\Sigma}$ is invertible as the rank of $\mathbf{M}$ is larger than p. Then for any orthogonal matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, $\hat{\mathbf{X}} = \mathbf{U}(:, 1 : p)\hat{\Sigma}\mathbf{A}$ minimizes the objective function $\frac{1}{2}\|\mathbf{X}\mathbf{X}^\top - \mathbf{M}\mathbf{M}^\top\|_F^2$.

After we find $\hat{\mathbf{X}} = \mathbf{U}(:,1:p)\hat{\Sigma}\mathbf{A}$ for a certain orthogonal matrix $\mathbf{A}$, we have

$$\begin{aligned}\mathbf{Y} = \mathbf{M}^\top\hat{\mathbf{X}}(\hat{\mathbf{X}}^\top\hat{\mathbf{X}})^{-1} =& \mathbf{V}\Sigma\mathbf{U}^\top\mathbf{U}(:,1:p)\hat{\Sigma}\mathbf{A}((\mathbf{U}(:,1:p)\hat{\Sigma}\mathbf{A})^\top\mathbf{U}(:,1:p)\hat{\Sigma}\mathbf{A})^{-1}\\ =&\mathbf{V}\Sigma\mathbf{U}^\top\mathbf{U}(:,1:p)\hat{\Sigma}^{-1}\mathbf{A}\\ =&\mathbf{V}(:,1:p)\hat{\Sigma}\hat{\Sigma}^{-1}\mathbf{A} = \mathbf{V}(:,1:p)\mathbf{A},\end{aligned}$$

where the third equality is due to the fact that

$$\Sigma\mathbf{U}^\top\mathbf{U}(:,1:p) = \begin{bmatrix} \hat{\Sigma}_{p\times p} \\ \mathbf{0}_{(q-p)\times p}\end{bmatrix}.$$

According to Theorem 2.2, $(\hat{\mathbf{X}}, \mathbf{Y})$ is an optimal solution of problem (10) if $\mu = 0$. Note that $\hat{\mathbf{X}} = \mathbf{U}(:,1:p)\hat{\Sigma}\mathbf{A}$ is the thin SVD with $\hat{\Sigma} \in \mathbb{R}^{p\times p}$. Then, the third step gives $\mathbf{X} = \mathbf{U}(:,1:p)\hat{\Sigma}_\lambda\mathbf{A}$. Theorem 2.2 shows that $(\mathbf{X}, \mathbf{Y})$ is an optimal solution of problem (10). □

Remark: To find $\hat{\mathbf{X}}$ in the first step, we apply the Gauss-Newton algorithm from [17], which is previously used for RPCA in [20]. The iteration is $\mathbf{X} \leftarrow \mathbf{M}\mathbf{M}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} - \mathbf{X}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{M}\mathbf{M}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} - \mathbf{I})/2$. When p is small, computing the inverse of $\mathbf{X}^\top\mathbf{X}$ is fast. Though an iterative algorithm is required to solve this subproblem at each outer iteration, we can use the output from the previous outer iteration as the initial and the number of inner iterations is reduced significantly. Therefore, the computational time can be reduced significantly, as shown in Section 3. In the numerical experiments, the first Gauss-Newton algorithm requires several hundred iterations, while the number for following Gauss-Newton algorithms reduces to less than ten.

From Theorem 2.2, we say that we solve the proximal operator of the nonconvex function $\|\mathbf{L}\|_* + \iota_{\mathrm{rank}(\mathbf{L})\le p}(\mathbf{L})$ exactly. Here the indicator function is defined as

$$\iota_{\mathrm{rank}(\mathbf{L})\le p}(\mathbf{L}) = \begin{cases} 0, & \text{if } \mathrm{rank}(\mathbf{L}) \le p;\\ +\infty, & \text{otherwise.}\end{cases}$$

With these theorems, we are ready to develop optimization algorithms for the general problem (7).

2.1. Forward-backward.

First, we eliminate $\mathbf{S}$, and it becomes the following problem with $\mathbf{L}$ only:

$$\begin{aligned}&\operatorname*{minimize}_{\mathbf{L}:\mathrm{rank}(\mathbf{L})\le p}\ \min_{\mathbf{S}} \frac{1}{2}\|\mathcal{A}(\mathbf{L}) + \mathbf{S} - \mathbf{D}\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mu\|\mathbf{L}\|_*\\ =&\operatorname*{minimize}_{\mathbf{L}:\mathrm{rank}(\mathbf{L})\le p}\ \min_{\mathbf{S}} \left\{\frac{1}{2}\|\mathcal{A}(\mathbf{L}) + \mathbf{S} - \mathbf{D}\|_F^2 + \lambda\|\mathbf{S}\|_1\right\} + \mu\|\mathbf{L}\|_*\\ =&\operatorname*{minimize}_{\mathbf{L}:\mathrm{rank}(\mathbf{L})\le p}\ f_\lambda(\mathbf{D} - \mathcal{A}(\mathbf{L})) + \mu\|\mathbf{L}\|_*.\end{aligned} \tag{12}$$

Here f_λ is the Moreau envelope of $\lambda|\cdot|$ defined by $f_\lambda(x) = \min_{y\in\mathbb{R}}\{\lambda|y| + \frac{1}{2}(y-x)^2\}$. So it is differential and has a 1-Lipschitz continuous gradient. Then we can apply the proximal-gradient method (or forward-backward operator splitting). We take the gradient of f_λ, which is given by

$$f'_\lambda(x) = x - \mathrm{sign}(x)\max(0, |x| - \lambda) = \mathrm{sign}(x)\min(\lambda, |x|). \tag{13}$$

The forward-backward iteration for $\mathbf{L}$ with stepsize t is

$$\mathbf{L}^{k+1} = \mathbf{prox}_{t\mu}\left(\mathbf{L}^k - t\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\right), \tag{14}$$

where the proximal operator is defined by

$$\mathbf{prox}_{\mu}(\mathbf{A}) = \mathop{\arg\min}_{\mathbf{L}:\mathrm{rank}(\mathbf{L})\leq p} \frac{1}{2}\|\mathbf{L}-\mathbf{A}\|_F^2 + \mu\|\mathbf{L}\|_*. \tag{15}$$

The algorithm is summarized in Alg. 1.

```
Algorithm 1: Proposed Algorithm
Input: $\mathbf{D}$, $\mu$, $\lambda$, $p$, $\mathcal{A}$, stepsize $t$, stopping criteria $\epsilon$, maximum number of
       iterations $Max\_Iter$, initialization $\mathbf{L}^0 = \mathbf{0}$
Output: $\mathbf{L}$, $\mathbf{S}$
for $k$ = 0, 1, 2, 3, ..., $Max\_Iter$ do
    $\mathbf{S} = \mathrm{sign}(\mathbf{D} - \mathcal{A}(\mathbf{L}^k)) \odot \max(0, |\mathbf{D} - \mathcal{A}(\mathbf{L}^k)| - \lambda)$ ;
    $\mathbf{L}^{k+1} = \mathbf{prox}_{t\mu}(\mathbf{L}^k - t\mathcal{A}^*(\mathcal{A}(\mathbf{L}^k) - \mathbf{D} + \mathbf{S})$ using Gauss-Newton;
    if $\|\mathbf{L}^{k+1} - \mathbf{L}^k\|_F / \|\mathbf{L}^k\|_F < \epsilon$ then
        break
    end
end
```

Connection to [20]**.** Consider the special case with $\mathcal{A} = \mathcal{I}$ and $\mu = 0$. We let $t = 1$ in (14) and obtain the following iteration

$$\mathbf{L}^{k+1} = \mathbf{prox}_0(\mathbf{L}^k - f'_\lambda(\mathbf{L}^k - \mathbf{D})) = \mathop{\arg\min}_{\mathbf{L}:\mathrm{rank}(\mathbf{L})\leq p} \frac{1}{2}\|\mathbf{L} + \mathbf{S}^{k+1} - \mathbf{D}\|^2,$$

where $\mathbf{S}^{k+1} = \mathrm{sign}(\mathbf{D} - \mathbf{L}^k) \odot \max(0, |\mathbf{D} - \mathbf{L}^k| - \lambda)$. This is exactly the algorithm in [20] for solving (5). It alternates between finding the best $\mathbf{S}$ with $\mathbf{L}$ fixed and the best $\mathbf{L}$ (or $(\mathbf{X}, \mathbf{Y})$) with $\mathbf{S}$ fixed.

Recently, the work [3] proposed a novel RPCA algorithm with linear convergence. It projects matrices to special manifolds of low-rank matrices, and their truncated SVD can be computed efficiently. Our matrix does not have this property in our algorithm, and a good initial guess from the previous iteration is necessary to reduce the computation in the Gauss-Newton method.

2.1.1. *Convergence analysis.* From the discussion above, problem (7) can be solved by an iteration process of forward-backward splitting. In each iteration, we reduce the value of the objective function

$$E(\mathbf{L}, \mathbf{S}) = \frac{1}{2}\|\mathcal{A}(\mathbf{L}) + \mathbf{S} - \mathbf{D}\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mu\|\mathbf{L}\|_* \tag{16}$$

by applying proximal operators to $\mathbf{L}$ and $\mathbf{S}$ alternatively. The resulting iteration sequence $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$ with some initial $(\mathbf{L}^0, \mathbf{S}^0)$ is explicitly given by

$$\begin{aligned} \mathbf{S}^k &= \mathrm{sign}(\mathbf{D} - \mathcal{A}(\mathbf{L}^{k-1})) \odot \max(0, |\mathbf{D} - \mathcal{A}(\mathbf{L}^{k-1})| - \lambda), \\ \mathbf{L}^k &= \mathbf{prox}_{t\mu}\left(\mathbf{L}^{k-1} - t\mathcal{A}^*(\mathcal{A}(\mathbf{L}^{k-1}) + \mathbf{S}^k - \mathbf{D})\right), \end{aligned} \tag{17}$$

where the proximal operator $\mathbf{prox}_{t\mu}(\cdot)$ for updating $\mathbf{L}$ is defined by (15). Here we use (13) to derive

$$\begin{aligned} & f'_\lambda(\mathcal{A}(\mathbf{L}^{k-1}) - \mathbf{D}) \\ &= \mathcal{A}(\mathbf{L}^{k-1}) - \mathbf{D} + \mathrm{sign}(\mathbf{D} - \mathcal{A}(\mathbf{L}^{k-1})) \odot \max(0, |\mathbf{D} - \mathcal{A}(\mathbf{L}^{k-1})| - \lambda) \\ &= \mathcal{A}(\mathbf{L}^{k-1}) + \mathbf{S}^k - \mathbf{D}. \end{aligned}$$

In this subsection, we establish the convergence results for $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$. We will show that every limit point of $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$, denoted by $(\mathbf{L}^\star, \mathbf{S}^\star)$, is a fixed point of the proximal operator, i.e.,

$$\begin{aligned} &\mathbf{S}^\star = \operatorname{sign}(\mathbf{D} - \mathcal{A}(\mathbf{L}^\star)) \odot \max(0, |\mathbf{D} - \mathcal{A}(\mathbf{L}^\star)| - \lambda), \\ &\mathbf{L}^\star = \mathbf{prox}_{t\mu}\left(\mathbf{L}^\star - t\mathcal{A}^*(\mathcal{A}(\mathbf{L}^\star) + \mathbf{S}^\star - \mathbf{D})\right). \end{aligned} \tag{18}$$

In practical execution, one can efficiently solve the proximal operator for $\mathbf{L}$ by solving $(\mathbf{X}^k, \mathbf{Y}^k)$ through

$$\begin{aligned} &\underset{\mathbf{X},\mathbf{Y}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{X}\mathbf{Y}^\top - \mathbf{L}^{k-1} + t\mathcal{A}^*(\mathcal{A}(\mathbf{L}^{k-1}) + \mathbf{S}^k - \mathbf{D})\|_F^2 + \mu\|\mathbf{X}\|_*, \\ &\text{subject to } \mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_{p\times p}, \end{aligned} \tag{19}$$

and letting $\mathbf{L}^k = \mathbf{X}^k(\mathbf{Y}^k)^\top$. We also prove that if $(\mathbf{X}^\star, \mathbf{Y}^\star, \mathbf{S}^\star)$ is a limit point of $\{(\mathbf{X}^k, \mathbf{Y}^k, \mathbf{S}^k)\}_{k\geq 1}$, then $(\mathbf{X}^\star(\mathbf{Y}^\star)^\top, \mathbf{S}^\star)$ is a limit point of $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$, and the limit point $(\mathbf{X}^\star, \mathbf{Y}^\star, \mathbf{S}^\star)$ is a stationary point of

$$E(\mathbf{X}\mathbf{Y}^\top, \mathbf{S}) = \frac{1}{2}\|\mathcal{A}(\mathbf{X}\mathbf{Y}^\top) + \mathbf{S} - \mathbf{D}\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mu\|\mathbf{X}\mathbf{Y}^\top\|_*,$$

i.e., $(\mathbf{X}^\star, \mathbf{Y}^\star, \mathbf{S}^\star)$ satisfies the first-order optimality condition

$$\begin{aligned} &\mathbf{0} \in [\mathcal{A}^*(\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) + \mathbf{S}^\star - \mathbf{D}) + \mu\partial\|\mathbf{X}^\star(\mathbf{Y}^\star)^\top\|_*]\mathbf{Y}^\star, \\ &\mathbf{0} \in (\mathbf{X}^\star)^\top[\mathcal{A}^*(\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) + \mathbf{S}^\star - \mathbf{D}) + \mu\partial\|\mathbf{X}^\star(\mathbf{Y}^\star)^\top\|_*], \\ &\mathbf{0} \in \mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) + \mathbf{S}^\star - \mathbf{D} + \lambda\partial\|\mathbf{S}^\star\|_1. \end{aligned} \tag{20}$$

We summarize these results in the following theorem.

Theorem 2.4. *Define the objective function $E(\mathbf{L}, \mathbf{S})$ as (16). Let $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$ be a sequence generated by (17) with initial $(\mathbf{L}^0, \mathbf{S}^0)$ and stepsize $t < \frac{1}{\|\mathcal{A}\|^2}$, where $\mathbf{L}^k = \mathbf{X}^k(\mathbf{Y}^k)^\top$ with $(\mathbf{X}^k, \mathbf{Y}^k)$ being solved from (19). We have the following statements:*

1. *The objective values $\{E(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$ are non-increasing along $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$.*
2. *The sequence $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$ is bounded and thus has limit points.*
3. *Every limit point $(\mathbf{L}^\star, \mathbf{S}^\star)$ of $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$ satisfies (18).*
4. *The sequence $\{(\mathbf{X}^k, \mathbf{Y}^k, \mathbf{S}^k)\}_{k\geq 1}$ is also bounded. In addition, for any limit point $(\mathbf{X}^\star, \mathbf{Y}^\star, \mathbf{S}^\star)$ of $\{(\mathbf{X}^k, \mathbf{Y}^k, \mathbf{S}^k)\}_{k\geq 1}$, $(\mathbf{X}^\star(\mathbf{Y}^\star)^\top, \mathbf{S}^\star)$ is a limit point of $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\geq 1}$.*
5. *Every limit point $(\mathbf{X}^\star, \mathbf{Y}^\star, \mathbf{S}^\star)$ of $\{(\mathbf{X}^k, \mathbf{Y}^k, \mathbf{S}^k)\}_{k\geq 1}$ is a stationary point of $E(\mathbf{X}\mathbf{Y}^\top, \mathbf{S})$, which satisfies the first-order optimality condition in (20).*

In addition, if $\mathcal{A} = \mathcal{I}$, we can take the stepsize $t = 1$, and all the statements above still hold.

Proof. We start by verifying the first two statements. For $k \geq 0$ and $t < \frac{1}{\|\mathcal{A}\|^2}$, we have

$$\begin{aligned}
&E(\mathbf{L}^{k+1}, \mathbf{S}^{k+1}) \\
&= \frac{1}{2}\|\mathcal{A}(\mathbf{L}^{k+1}) - \mathcal{A}(\mathbf{L}^k)\|_F^2 + \langle \mathcal{A}(\mathbf{L}^{k+1}) - \mathcal{A}(\mathbf{L}^k), \mathcal{A}(\mathbf{L}^k) + \mathbf{S}^{k+1} - \mathbf{D}\rangle \\
&\quad + \frac{1}{2}\|\mathcal{A}(\mathbf{L}^k) + \mathbf{S}^{k+1} - \mathbf{D}\|_F^2 + \lambda\|\mathbf{S}^{k+1}\|_1 + \mu\|\mathbf{L}^{k+1}\|_* \\
&\leq \frac{1}{2t}\|\mathbf{L}^{k+1} - \mathbf{L}^k\|_F^2 + \langle \mathbf{L}^{k+1} - \mathbf{L}^k, \mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\rangle + \mu\|\mathbf{L}^{k+1}\|_* \\
&\quad + \frac{1}{2}\|\mathcal{A}(\mathbf{L}^k) + \mathbf{S}^{k+1} - \mathbf{D}\|_F^2 + \lambda\|\mathbf{S}^{k+1}\|_1 + \left(\frac{\|\mathcal{A}\|^2}{2} - \frac{1}{2t}\right)\|\mathbf{L}^{k+1} - \mathbf{L}^k\|_F^2 \\
&= \frac{1}{t}\left\{\frac{1}{2}\|\mathbf{L}^{k+1} - \mathbf{L}^k + t\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\|_F^2 + t\mu\|\mathbf{L}^{k+1}\|_*\right\} \\
&\quad - \frac{t}{2}\|\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\|_F^2 + \left(\frac{\|\mathcal{A}\|^2}{2} - \frac{1}{2t}\right)\|\mathbf{L}^{k+1} - \mathbf{L}^k\|_F^2 \\
&\quad + \frac{1}{2}\|\mathcal{A}(\mathbf{L}^k) + \mathbf{S}^{k+1} - \mathbf{D}\|_F^2 + \lambda\|\mathbf{S}^{k+1}\|_1,
\end{aligned} \tag{21}$$

where the inequality is due to the facts that

$$\|\mathcal{A}(\mathbf{L}^{k+1}) - \mathcal{A}(\mathbf{L}^k)\|_F^2 \leq \|\mathcal{A}\|^2\|\mathbf{L}^{k+1} - \mathbf{L}^k\|_F^2$$

and

$$\mathcal{A}(\mathbf{L}^k) + \mathbf{S}^{k+1} - \mathbf{D} = f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D}).$$

Note that $\mathbf{L}^{k+1} = \mathbf{prox}_{t\mu}\left(\mathbf{L}^k - t\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\right)$, which solves

$$\underset{\mathbf{L}:\mathrm{rank}(\mathbf{L})\leq p}{\text{minimize}} \ \frac{1}{2}\|\mathbf{L} - \mathbf{L}^k + t\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\|_F^2 + t\mu\|\mathbf{L}\|_*.$$

Since $\mathrm{rank}(\mathbf{L}^k) \leq p$, we have

$$\begin{aligned}
&\frac{1}{2}\|\mathbf{L}^{k+1} - \mathbf{L}^k + t\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\|_F^2 + t\mu\|\mathbf{L}^{k+1}\|_* \\
&\leq \frac{1}{2}\|\mathbf{L}^k - \mathbf{L}^k + t\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\|_F^2 + t\mu\|\mathbf{L}^k\|_* \\
&= \frac{t^2}{2}\|\mathcal{A}^* f'_\lambda(\mathcal{A}(\mathbf{L}^k) - \mathbf{D})\|_F^2 + t\mu\|\mathbf{L}^k\|_*.
\end{aligned}$$

Substituting the above estimate to (21) yields

$$\begin{aligned}
E(\mathbf{L}^{k+1}, \mathbf{S}^{k+1}) \leq &\left(\frac{\|\mathcal{A}\|^2}{2} - \frac{1}{2t}\right)\|\mathbf{L}^{k+1} - \mathbf{L}^k\|_F^2 \\
&+ \frac{1}{2}\|\mathcal{A}(\mathbf{L}^k) + \mathbf{S}^{k+1} - \mathbf{D}\|_F^2 + \mu\|\mathbf{L}^k\|_* + \lambda\|\mathbf{S}^{k+1}\|_1.
\end{aligned} \tag{22}$$

Moreover, we see that

$$\mathbf{S}^{k+1} = \arg\min_{\mathbf{S}} \ \frac{1}{2}\|\mathbf{S} - (\mathbf{D} - \mathcal{A}(\mathbf{L}^k))\|_F^2 + \lambda\|\mathbf{S}\|_1.$$

Then from [18, Lemma 2], there holds

$$\begin{aligned}(23)\qquad &\frac{1}{2}\|\mathbf{S}^{k+1}-(\mathbf{D}-\mathcal{A}(\mathbf{L}^k))\|_F^2+\lambda\|\mathbf{S}^{k+1}\|_1\\ &\le\frac{1}{2}\|\mathbf{S}^{k}-(\mathbf{D}-\mathcal{A}(\mathbf{L}^k))\|_F^2+\lambda\|\mathbf{S}^{k}\|_1-\frac{1}{2}\|\mathbf{S}^{k+1}-\mathbf{S}^k\|_F^2.\end{aligned}$$

Combining estimates (22) and (23), we find that

$$(24)\quad E(\mathbf{L}^{k+1},\mathbf{S}^{k+1})\le E(\mathbf{L}^k,\mathbf{S}^k)+\left(\frac{\|\mathcal{A}\|^2}{2}-\frac{1}{2t}\right)\|\mathbf{L}^{k+1}-\mathbf{L}^k\|_F^2-\frac{1}{2}\|\mathbf{S}^{k+1}-\mathbf{S}^k\|_F^2.$$

Since $\frac{\|\mathcal{A}\|^2}{2}-\frac{1}{2t}<0$, the estimate above implies $E(\mathbf{L}^{k+1},\mathbf{S}^{k+1})\le E(\mathbf{L}^k,\mathbf{S}^k)$ for any $k\ge 0$, which verifies the first statement.

Note that the target function $E(\mathbf{L},\mathbf{S})$ is coercive, i.e., $E(\mathbf{L},\mathbf{S})\to+\infty$ when $\|\mathbf{L}\|_F+\|\mathbf{S}\|_F\to+\infty$. Since $E(\mathbf{L}^k,\mathbf{S}^k)\le E(\mathbf{L}^0,\mathbf{S}^0)<+\infty,\forall k\ge 1$, this property guarantees that both $\{\mathbf{L}^k\}_{k\ge1}$ and $\{\mathbf{S}^k\}_{k\ge1}$ are bounded sequences, and thus the second statement holds.

For any limit point $(\mathbf{L}^\star,\mathbf{S}^\star)$ of $\{(\mathbf{L}^k,\mathbf{S}^k)\}_{k\ge1}$, there exists a convergent subsequence $\{(\mathbf{L}^{k_i},\mathbf{S}^{k_i})\}_{i\ge1}$ such that $\mathbf{L}^{k_i}\to\mathbf{L}^\star$ and $\mathbf{S}^{k_i}\to\mathbf{S}^\star$. On the other hand, we see that

$$\begin{aligned}(25)\qquad &\mathbf{S}^{k_i+1}=\operatorname{sign}(\mathbf{D}-\mathcal{A}(\mathbf{L}^{k_i}))\odot\max(0,|\mathbf{D}-\mathcal{A}(\mathbf{L}^{k_i})|-\lambda),\\ &\mathbf{L}^{k_i+1}=\mathbf{prox}_{t\mu}\left(\mathbf{L}^{k_i}-t\mathcal{A}^*(\mathcal{A}(\mathbf{L}^{k_i})+\mathbf{S}^{k_i+1}-\mathbf{D})\right).\end{aligned}$$

Summing both sides of (24) from $k=0$ to ∞, we obtain

$$\left(\frac{1}{t}-\|\mathcal{A}\|^2\right)\sum_{k=0}^{\infty}\|\mathbf{L}^{k+1}-\mathbf{L}^k\|_F^2+\sum_{k=0}^{\infty}\|\mathbf{S}^{k+1}-\mathbf{S}^k\|_F^2\le 2E(\mathbf{L}^0,\mathbf{S}^0)<\infty.$$

This inequality guarantees that $\{\mathbf{S}^{k_i+1}\}_{i\ge1}$ has the same limit point $\mathbf{S}^\star$ as that of $\{\mathbf{S}^{k_i}\}_{i\ge1}$, and $\{\mathbf{L}^{k_i+1}\}_{i\ge1}$ has the same limit point $\mathbf{L}^\star$ as that of $\{\mathbf{L}^{k_i}\}_{i\ge1}$. Then by taking limits in both sides of the two equations in (25), we obtain the third statement.

Next we will prove the last two statements. As $\|\mathbf{X}^k\|_F^2=\|\mathbf{L}^k\|_F^2$ and $\|\mathbf{Y}^k\|_F^2=p$, we know that the sequence $\{(\mathbf{X}^k,\mathbf{Y}^k,\mathbf{S}^k)\}_{k\ge1}$ is also bounded. Let $(\mathbf{X}^\star,\mathbf{Y}^\star,\mathbf{S}^\star)$ be a limit point of $\{(\mathbf{X}^k,\mathbf{Y}^k,\mathbf{S}^k)\}_{k\ge1}$, which is the limitation of a subsequence $\{(\mathbf{X}^{k_i},\mathbf{Y}^{k_i},\mathbf{S}^{k_i})\}_{i\ge1}$. Then we have

$$\mathbf{L}^{k_i}=\mathbf{X}^{k_i}(\mathbf{Y}^{k_i})^\top\to\mathbf{X}^\star(\mathbf{Y}^\star)^\top\text{ and }\mathbf{S}^{k_i}\to\mathbf{S}^\star,$$

i.e., $(\mathbf{X}^\star(\mathbf{Y}^\star)^\top,\mathbf{S}^\star)$ is the limit point of $\{(\mathbf{L}^k,\mathbf{S}^k)\}_{k\ge1}$ achieved by the subsequence $\{(\mathbf{L}^{k_i},\mathbf{S}^{k_i})\}_{i\ge1}$. Thus the fourth statement is verified.

Now we are in the position to prove the fifth statement. Due to the third and fourth statements, if $(\mathbf{X}^\star,\mathbf{Y}^\star,\mathbf{S}^\star)$ is a limit point of $\{(\mathbf{X}^k,\mathbf{Y}^k,\mathbf{S}^k)\}_{k\ge1}$, i.e., $(\mathbf{X}^\star(\mathbf{Y}^\star)^\top,\mathbf{S}^\star)$ should satisfy (18)

$$\begin{aligned}(26)\qquad &\mathbf{S}^\star=\operatorname{sign}(\mathbf{D}-\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top))\odot\max(0,|\mathbf{D}-\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top)|-\lambda),\\ &\mathbf{X}^\star(\mathbf{Y}^\star)^\top=\mathbf{prox}_{t\mu}\left(\mathbf{X}^\star(\mathbf{Y}^\star)^\top-t\mathcal{A}^*(\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top)+\mathbf{S}^\star-\mathbf{D})\right).\end{aligned}$$

The first condition in (26) implies that the limit point $\mathbf{S}^\star$ minimizes

$$\frac{1}{2}\|\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top)+\mathbf{S}-\mathbf{D}\|_F^2+\lambda\|\mathbf{S}\|_1+\mu\|\mathbf{X}^\star(\mathbf{Y}^\star)^\top\|_*$$

over all $\mathbf{S}\in\mathbb{R}^{m\times n}$. Thus, $\mathbf{S}^\star$ should satisfy the third condition in (20).

Moreover, since $\operatorname{rank}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) \le p$, the second condition in (26) actually implies that $(\mathbf{X}^\star, \mathbf{Y}^\star)$ is an optimal solution of the problem

$$\operatorname*{minimize}_{\mathbf{X},\mathbf{Y}} \ \frac{1}{2}\|\mathbf{X}\mathbf{Y}^\top - \mathbf{X}^\star(\mathbf{Y}^\star)^\top + t\mathcal{A}^*(\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) + \mathbf{S}^\star - \mathbf{D})\|_F^2 + t\mu\|\mathbf{X}\mathbf{Y}^\top\|_*.$$

Therefore, $(\mathbf{X}^\star, \mathbf{Y}^\star)$ should satisfy the first-order optimality condition for $\mathbf{X}$, which gives

$$\begin{aligned}&[\mathbf{X}^\star(\mathbf{Y}^\star)^\top - \mathbf{X}^\star(\mathbf{Y}^\star)^\top + t\mathcal{A}^*(\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) + \mathbf{S}^\star - \mathbf{D})]\mathbf{Y}^\star \\ &\quad + t\mu\partial\|\mathbf{X}^\star(\mathbf{X}^\star(\mathbf{Y}^\star)^\top)^\top\|_*\mathbf{Y}^\star \\ =&t[\mathcal{A}^*(\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) + \mathbf{S}^\star - \mathbf{D}) + \mu\partial\|\mathbf{X}^\star(\mathbf{Y}^\star)^\top\|_*]\mathbf{Y}^\star \ni \mathbf{0}.\end{aligned}$$

Similarly, from the first-order opitmality condition for $\mathbf{Y}$, one can verify that

$$\mathbf{0} \in (\mathbf{X}^\star)^\top[\mathcal{A}^*(\mathcal{A}(\mathbf{X}^\star(\mathbf{Y}^\star)^\top) + \mathbf{S}^\star - \mathbf{D}) + \mu\partial\|\mathbf{X}^\star(\mathbf{Y}^\star)^\top\|_*].$$

We thus derive the first two conditions in (20).

We will complete our proof by verifying the convergence results for the special case of $\mathcal{A} = \mathcal{I}$ and $t = 1$. In this case, by the same method, one can derive a similar inequality as (24), which is

$$E(\mathbf{L}^{k+1}, \mathbf{S}^{k+1}) \le E(\mathbf{L}^k, \mathbf{S}^k) - \frac{1}{2}\|\mathbf{S}^{k+1} - \mathbf{S}^k\|_F^2.$$

Then $\{E(\mathbf{L}^k, \mathbf{S}^k)\}_{k\ge1}$ are non-increasing along $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\ge1}$, and $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\ge1}$ is bounded due to the coerciveness of $E(\mathbf{L}, \mathbf{S})$. Let $(\mathbf{L}^\star, \mathbf{S}^\star)$ be the limit point of $\{(\mathbf{L}^k, \mathbf{S}^k)\}_{k\ge1}$ achieved by the subsequence $\{(\mathbf{L}^{k_i}, \mathbf{S}^{k_i})\}_{i\ge1}$. Recall the iterations for updating $\mathbf{S}^{k_i+1}$ and $\mathbf{L}^{k_i}$ given by

$$\begin{aligned}\mathbf{S}^{k_i+1} &= \operatorname{sign}(\mathbf{D} - \mathbf{L}^{k_i}) \odot \max(0, |\mathbf{D} - \mathbf{L}^{k_i}| - \lambda), \\ \mathbf{L}^{k_i} &= \mathbf{prox}_\mu\left(\mathbf{D} - \mathbf{S}^{k_i}\right).\end{aligned} \tag{27}$$

Since $\sum_{k=0}^\infty \|\mathbf{S}^{k+1} - \mathbf{S}^k\|_F^2 \le 2E(\mathbf{L}^0, \mathbf{S}^0) < +\infty$, $\{\mathbf{S}^{k_i+1}\}_{i\ge1}$ has the same limit point $\mathbf{S}^\star$ as that of $\{\mathbf{S}^{k_i}\}_{i\ge1}$. Taking limits in both sides of equations (27) yields the condition (18) for $\mathcal{A} = \mathcal{I}$ and $t = 1$. The last two statements can be verified by exactly the same arguments for the general case. We thus complete the proof. □

2.2. An accelerated algorithm.

We show in the previous subsection that Alg. 1 is a forward-backward splitting or proximal gradient algorithm for a nonconvex problem. Recently, accelerated proximal gradient (APG) algorithms are proposed for nonconvex problems to reduce the computational time without sacrificing convergence [13, 14]. In this paper, we adopt the nonmonotone APG [14, Alg. 2] because of its better performance shown in [14]. The algorithm is described in Alg. 2. We let $\delta = 1$ and $\eta = 0.6$ in the numerical experiments.

3. Numerical experiments.

In this section, we use synthetic data and real images to demonstrate the performance of our proposed model and algorithms. The code to reproduce the results in this section can be found at https://github.com/mingyan08/RPCA_Rank_Bound.

```
Algorithm 2: Accelerated algorithm with nonmonotone APG
  Input: $\mathbf{D}$, $\mu$, $\lambda$, $p$, $\mathcal{A}$, stepsize $t$, $\eta \in [0,1)$, $\delta > 0$, stopping criteria $\epsilon$,
         maximum number of iterations $Max\_Iter$, initialization:
         $\mathbf{L}^0 = \mathbf{L}^1 = \mathbf{Z}^1 = \mathbf{0}$, $t^0 = 0$, $t^1 = q^1 = 1$, $c^1 = F(\mathbf{L}^1)$
  Output: $\mathbf{L}$, $\mathbf{S}$
  for $k$ = 1, 2, 3, .., $Max\_Iter$ do
      $\mathbf{L} = \mathbf{L}^k + \frac{t^{k-1}}{t^k}(\mathbf{Z}^k - \mathbf{L}^k) + \frac{t^{k-1}-1}{t^k}(\mathbf{L}^k - \mathbf{L}^{k-1})$;
      $\mathbf{S} = \mathrm{sign}(\mathbf{D} - \mathcal{A}(\mathbf{L})) \odot \max(0, |\mathbf{D} - \mathcal{A}(\mathbf{L})| - \lambda)$;
      $\mathbf{Z}^{k+1} = \mathbf{prox}_{t\mu}(\mathbf{L} - t\mathcal{A}^*(\mathcal{A}(\mathbf{L}) - \mathbf{D} + \mathbf{S}))$;
      if $F(\mathbf{Z}^{k+1}) \leq c^k - \delta\|\mathbf{Z}^{k+1} - \mathbf{L}\|^2$ then
          $\mathbf{L}^{k+1} = \mathbf{Z}^{k+1}$;
      else
          $\mathbf{S}^k = \mathrm{sign}(\mathbf{D} - \mathcal{A}(\mathbf{L}^k)) \odot \max(0, |\mathbf{D} - \mathcal{A}(\mathbf{L}^k)| - \lambda)$;
          $\mathbf{V}^{k+1} = \mathbf{prox}_{t\mu}(\mathbf{L}^k - t\mathcal{A}^*(\mathcal{A}(\mathbf{L}^k) - \mathbf{D} + \mathbf{S}^k))$;
          $\mathbf{L}^{k+1} = \begin{cases} \mathbf{Z}^{k+1} & \textbf{if } F(\mathbf{Z}^{k+1}) \leq F(\mathbf{V}^{k+1}); \\ \mathbf{V}^{k+1} & \textbf{otherwise}; \end{cases}$
      end
      if $\|\mathbf{L}^k - \mathbf{L}^{k-1}\|_F / \|\mathbf{L}^{k-1}\|_F < \epsilon$ then
          break
      end
      $t^{k+1} = \frac{\sqrt{4(t^k)^2+1}+1}{2}$;
      $q^{k+1} = \eta q^k + 1$;
      $c^{k+1} = \frac{\eta q^k c^k + F(\mathbf{L}^{k+1})}{q^{k+1}}$;
  end
```

3.1. **Synthetic data.** We would like to recover the low-rank matrix from a noisy matrix that is contaminated by a sparse matrix and Gaussian noise. We create a true low-rank 500×500 matrix $\mathbf{L}^\star$ by multiplying a random $500 \times r$ matrix and a random $r \times 500$ matrix, where their components are generated from standard normal distribution independently. We calculate the mean of the absolute values of all the components in $\mathbf{L}^\star$ and denote it as c. Then we randomly select $s\%$ of the components and replace their values with uniformly distributed random values from $[-3c, 3c]$. After that, we add small Gaussian noise $\mathcal{N}(0, \sigma^2)$ to all components of the matrix. We let $t = 1.7$ in the experiments because of fast convergence, though the convergence results in Theorem 2.4 require $t < 1$.

3.1.1. *Low-rank matrix recovery.* We fix $\sigma = 0.05$ for the Gaussian noise and set the upper bound of the rank to be $p = r + 5$. We stop all algorithms when the relative error at the k-th iteration, which is defined as

$$RE(\mathbf{L}^{k+1}, \mathbf{L}^k) := \frac{\|\mathbf{L}^{k+1} - \mathbf{L}^k\|_F}{\|\mathbf{L}^k\|_F},$$

is less than 10^{-4}. We use the relative error to $\mathbf{L}^\star$, which is defined as

$$RE(\mathbf{L}, \mathbf{L}^*) := \frac{\|\mathbf{L} - \mathbf{L}^\star\|_F}{\|\mathbf{L}^\star\|_F},$$

to evaluate the performance of our proposed model and that in [20]. First, we consider the case with $r = 25$ and $s = 20$. We plot a contour map of the relative error to $\mathbf{L}^\star$ for different parameters μ and λ in Fig. 1. From this contour map, we can see that the best parameter does not happen when $\mu = 0$, which corresponds to the model in [20]. It verifies the better performance of our proposed model with appropriate parameters. In this subsection, we set $\lambda = 0.02$ for Shen et al.'s and $(\mu = 0.6,\ \lambda = 0.04)$ for our proposed algorithms.

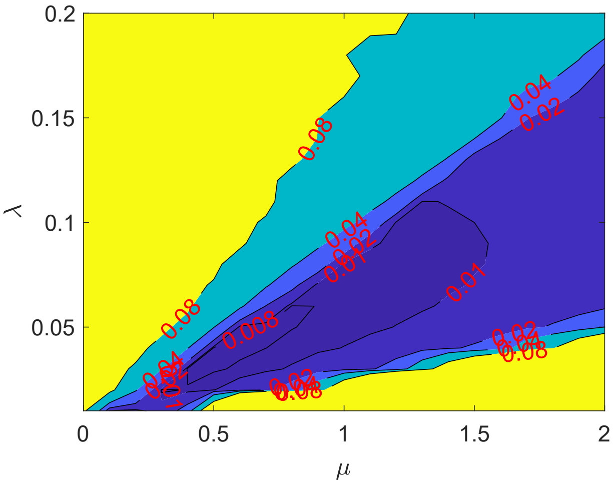


Figure 1. The contour map of the relative error to $\mathbf{L}^\star$ for different parameters. In this experiment, we set $r = 25$ and $s = 20$. The upper bound of the rank is set to be $p = 30$.

In addition, we consider another two settings for (r, s), and the comparison with different algorithms is shown in Table 1. In this table, we also compare the number of iterations for three algorithms: Shen et al.'s, Alg. 1, and Alg. 2. From this table, we can see that both Alg. 1 and Alg. 2 have better performance and fewer iterations than [20]. The accelerated Alg. 2 has the fewest iterations, but its performance in terms of $RE(\mathbf{L}, \mathbf{L}^\star)$ is not as good as Alg. 1 for the last case. It is because we stop both algorithms when the stopping criteria is satisfied, and the algorithms are not converged yet. We checked the objective function values for both algorithms, and the value for Alg. 2 is smaller than that for Alg. 1 in this case. Therefore, if we want a solution close to the true low-rank matrix $\mathbf{L}^\star$, we may need to stop early before the convergence, which is the same as many models for inverse problems.

3.1.2. *Robustness of the model.* In this experiment, we compare the robustness of our proposed model with that of [20]. We let $r = 25$ and $s = 20$. Then we run both models for p from 15 to 35. The comparison of the relative error to $\mathbf{L}^\star$ is shown in Fig. 2. We let $\lambda = 0.02$ for Shen et al.'s and $(\mu = 0.6,\ \lambda = 0.04)$ for Alg. 2. It shows that our proposed model is robust to the parameter p, as long as it is not smaller than the true rank r.

3.1.3. *Low-rank matrix recovery with missing entries.* In this experiment, we try to recover the low-rank matrix when there are missing entries in the matrix. Therefore,

r	s	Shen et al.'s [20]		Alg. 1		Alg.2	
		$RE(\mathbf{L},\mathbf{L}^\star)$	# iter	$RE(\mathbf{L},\mathbf{L}^\star)$	# iter	$RE(\mathbf{L},\mathbf{L}^\star)$	# iter
25	20	0.0745	1318	0.0075	296	0.0075	68
50	20	0.0496	1434	0.0101	473	0.0088	77
25	40	0.0990	2443	0.0635	796	0.0915	187

Table 1. Comparison of three RPCA algorithms. We compare the relative error of their solutions to the true low-rank matrix and the number of iterations. Both Alg. 1 and Alg. 2 have better performance than [20] in terms of the relative error and the number of iterations. Alg. 2 has the fewest iterations but the relative error could be large. It is because the true low-rank matrix is not the optimal solution to the optimization problem, and the trajectory of the iterations moves close to $\mathbf{L}^\star$ before it approaches the optimal solution.

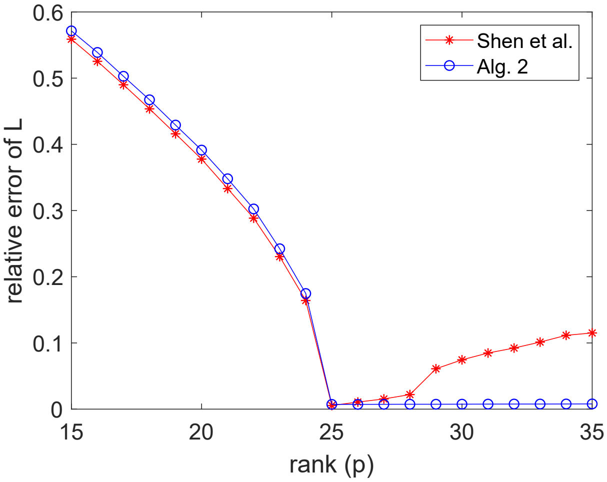


Figure 2. The relative error to the true low-rank matrix vs the rank p for Shen et al.'s and Alg. 2. Alg. 2 is robust to p, as long as p is not smaller than the true rank 25.

the operator $\mathcal{A}$ is not the identity $\mathcal{I}$. We randomly select the missing entries from all the entries. We let $r = 25$ and add both the sparse noise with parameter s and the Gaussian noise with parameter σ to the true matrix $\mathbf{L}^\star$. Then we apply Alg. 2 to recover the low-rank matrix, and the relative error to $\mathbf{L}^\star$ is used to evaluate the performance. The results for different settings are in Table 2. For the first three cases with $s = 20$, we choose $(\mu = 0.5,\ \lambda = 0.04)$, while we let $(\mu = 0.1,\ \lambda = 0.01)$ for the last case with $s = 5$. Note that, even with missing entries, Alg. 2 can reconstruct the low-rank matrix accurately.

3.2. **Real image experiment.** In this section, we consider the three algorithms applied to image processing problems. Since natural images are not low-rank essentially, we consider two cases on two different images ('cameraman' and 'Barbara').

s	σ	ratio of missing entries	$RE(\mathbf{L},\mathbf{L}^\star)$ by Alg. 2
20	0.05	10%	0.0079
20	0.05	20%	0.0088
20	0.05	50%	0.0201
5	0.01	50%	0.0015

Table 2. Performance of Alg. 2 on low-rank matrix recovery with missing entries. We change the level of sparsity in the sparse noise, standard deviation of the Gaussian noise, and the ratio of missing entries.

For the 256×256 cameraman image (the pixel values are from 0 to 255), we create an image with rank 37 from a low-rank approximation of the original image. Then we add 20% salt and pepper impulse noise and Gaussian noise with standard variance 4. We set 42 as the upper bound of the rank of the low-rank image for all algorithms. We let $\lambda = 0.03$ for Shen et al. and $(\mu = 0.5,\ \lambda = 0.06)$ for our model. To compare the performance of both models, we use the relative error defined in the last subsection and peak signal to noise ratio (PSNR) defined as

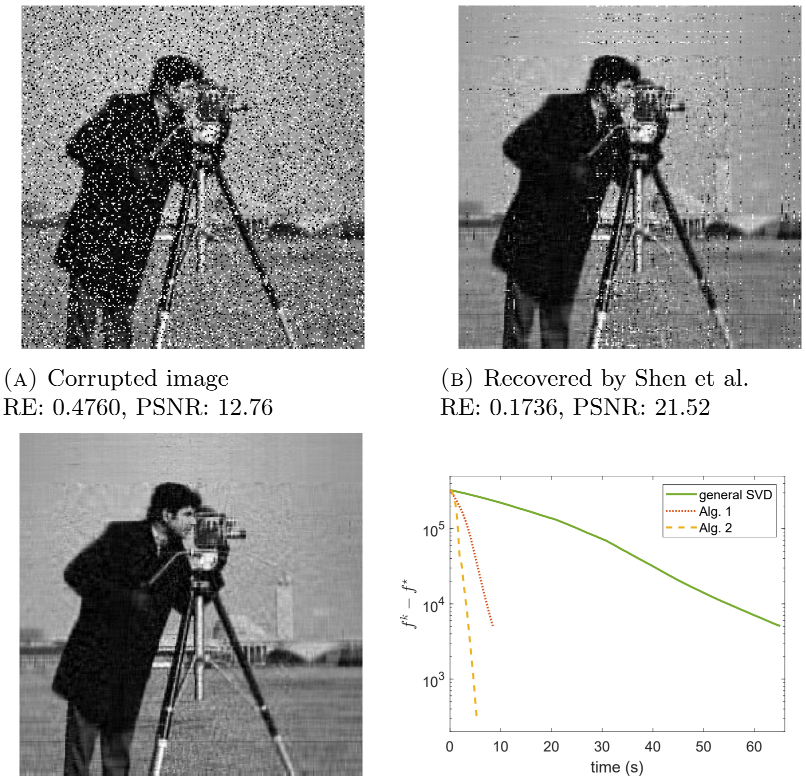
$$\text{PSNR} := 10 \log_{10} \frac{\text{Peak\_Val}^2}{\text{MSE}}.$$

Here Peak_Val is the largest value allowed at a pixel (255 in our case), and MSE is the mean squared error between the recovered image and the true image. The numerical results are shown in Fig. 3. From Fig. 3(A-C), we can see that our proposed model performs better than Shen et al. [20]. For the proposed model, we also compare the speed of three algorithms: Alg. 1, Alg. 1 with standard SVD, and Alg. 2 in Fig. 3(D). For both plots, we can see that the Gauss-Newton approach increases the speed comparing to the standard SVD approach. From the decrease of the objective function value, we can see that the accelerated algorithm Alg. 2 is faster than the nonaccelerated Alg. 1.

Next, we use the original 512×512 barbara image (the pixel values are from 0 to 255) without modification and add the same two types of noise as in the cameraman image. Because the original image is not low-rank, we choose the upper bound of rank $p = 50$. We let $\lambda = 0.03$ for Shen et al. and $(\mu = 0.5,\ \lambda = 0.06)$ for our model. The comparison result is shown in Fig. 4, and it is similar to the cameraman image. We also applied the acceleration to Shen et al.'s algorithm and obtained a better image with RE = 0.1447 and PSNR = 22.37.

4. Concluding remarks.

In this paper, we introduced a new model for RPCA when an upper bound of the rank is provided. For the unconstrained RPCA problem, we formulate it as the sum of one smooth function and one nonsmooth nonconvex function. Then we derive an algorithm based on proximal-gradient. This proposed algorithm has the alternating minimization algorithm [20] as a special case. Because of the connection between this algorithm and proximal gradient, we adopted an acceleration approach and proposed an accelerated algorithm. Both proposed algorithms have two advantages comparing to existing algorithms. First, different from algorithms that require accurate rank estimations, the proposed algorithms are robust to the upper bound of the rank. Second, we apply the Gauss-Newton algorithm to avoid the computation of singular values for large matrices,



(A) Corrupted image
RE: 0.4760, PSNR: 12.76

(B) Recovered by Shen et al.
RE: 0.1736, PSNR: 21.52

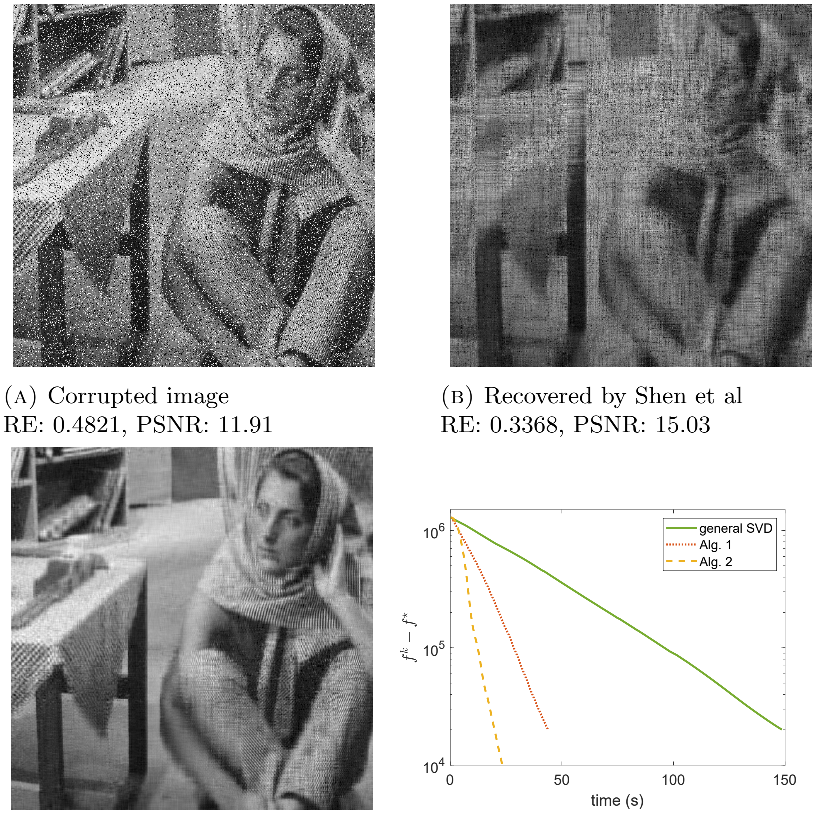
(C) Recovered by Alg. 2
RE: 0.0457, PSNR:33.11

(D) Comparison of the objective function value vs time for three algorithms

Figure 3. The numerical experiment on the 'cameraman' image. (A-C) show that the proposed model performs better than Shen et al.'s both visually and in terms of RE and PSNR. (D) compares the objective values vs time for general SVD, Alg. 1, and Alg. 2. Here $f^\star$ is the value obtained by Alg. 2 with more iterations. It shows the fast speed with the Gauss-Newton approach and acceleration. With the Gauss-Newton approach, the computation time for Alg. 1 is reduced to about 1/7 of the one with standard SVD (from 65.11s to 8.43s). The accelerated Alg. 2 requires 5.2s, though the number of iterations is reduced from 3194 to 360.

so our algorithm is faster than those algorithms that require SVD. Except for problem (7), this algorithm can be generalized to solve many other variants.

4.1. **Nonconvex penalties on the singular values.** In the problem (7), we choose the convex nuclear norm for the low-rank component in the objective function, which is the ℓ_1 norm on the singular values. The ℓ_1 norm pushes all singular values toward zero for the same amount, bringing bias in the solution. To promote the low-rankness of the low-rank component (or sparsity of its singular values), we can choose nonconvex regularization terms for the singular values. The idea for nonconvex regularization is to reduce the bias by pushing less on larger singular values. Some examples of nonconvex regularization are ℓ_p $(0 \leq p < 1)$ [5], smoothly



(A) Corrupted image
RE: 0.4821, PSNR: 11.91

(B) Recovered by Shen et al
RE: 0.3368, PSNR: 15.03

(C) Recovered by Alg. 2
RE: 0.1317, PSNR: 23.18

(D) Comparison of the objective function value vs time for three algorithms

FIGURE 4. The numerical experiment on the 'Barbara' image. (A-C) show that the proposed model performs better than Shen et al.'s both visually and in terms of RE and PSNR. (D) compares the objective values vs time for general SVD, Alg. 1, and Alg. 2. Here $f^\star$ is the value obtained by Alg. 2 with more iterations. It shows the fast speed with the Gauss-Newton approach and acceleration. With the Gauss-Newton approach, the computation time for Alg. 1 is reduced to less than 1/3 of the one with standard SVD (from 148.6s to 43.7s). The accelerated Alg. 2 requires 23.3s, though the number of iterations is reduced from 3210 to 300.

clipped absolute deviation (SCAD) [10], minimax concave penalty (MCP) [27], nonconvex weighted ℓ_1 [12], etc. When these regularization terms are applied, the only difference is in the third step for finding $\mathbf{X}$ in Lemma 2.3. Currently, we have to apply the soft thresholding on the singular values. When nonconvex regularization is used, we apply the corresponding thresholding on the singular values. In this case, all the convergence results stay valid.

4.2. **Other regularization on the sparse component.** We can also replace the ℓ_1 norm of the sparse component with other regularization terms. Similarly to the penalty on the singular values, the ℓ_1 norm on the sparse component brings bias, and we can use nonconvex regularization terms. Paper [23] uses both nonconvex regularization terms for the low-rank and sparse components. When different regularization terms are used on the sparse component, the new function f_λ (see (12) for the definition) may not be differentiable any more. In this case, the convergence results do not hold.

4.3. **Constrained problems.** When there is no noise in the measurements, the problem becomes constrained, and the previous algorithm can not be applied directly. Reference [20] uses the penalty method and gradually increases the weight for the penalization to approximate the constrained problem. Here, we introduce a new method based on ADMM. We consider the following constrained problem

$$\operatorname*{minimize}_{\mathbf{L},\mathbf{S}} \ \mu\|\mathbf{L}\|_* + \|\mathbf{S}\|_1, \ \text{subject to } \operatorname{rank}(\mathbf{L}) \le p, \ \mathbf{D} = \mathbf{L} + \mathbf{S}. \tag{28}$$

When we apply ADMM, the steps are

$$\mathbf{L}^{k+1} = \operatorname*{arg\,min}_{\mathbf{L}:\operatorname{rank}(\mathbf{L})\le p} \ \mu\|\mathbf{L}\|_* + \frac{\alpha}{2}\|\mathbf{D} - \mathbf{L} - \mathbf{S}^k + \frac{\mathbf{Z}^k}{\alpha}\|_F^2; \tag{29a}$$

$$\mathbf{S}^{k+1} = \operatorname*{arg\,min}_{\mathbf{S}} \|\mathbf{S}\|_1 + \frac{\alpha}{2}\|\mathbf{D} - \mathbf{L}^{k+1} - \mathbf{S} + \frac{\mathbf{Z}^k}{\alpha}\|_F^2; \tag{29b}$$

$$\mathbf{Z}^{k+1} = \mathbf{Z}^k - \alpha(\mathbf{L}^{k+1} + \mathbf{S}^{k+1} - D). \tag{29c}$$

The first step is exactly the proximal operator that can be solved from Lemma 2.3. The other two steps are easy to compute. This algorithm has only one parameter α, while penalty methods, such as that in [20], require additional parameters to increase the weight for the penalization.

Acknowledgments. The authors thank Dr. Yuan Shen for sharing the code of the algorithm proposed in [20]. The authors would like to thank two anonymous reviewers for their helpful comments and suggestions.

REFERENCES

[1] E. Amaldi and V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, *Theoretical Computer Science*, **209** (1998), 237–260.

[2] T. Bouwmans and E. H. Zahzah, Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance, *Computer Vision and Image Understanding*, **122** (2014), 22–34.

[3] H. Cai, J.-F. Cai and K. Wei, Accelerated alternating projections for robust principal component analysis, *The Journal of Machine Learning Research*, **20** (2019), 685–717.

[4] E. J. Candès, X. Li, Y. Ma and J. Wright, Robust principal component analysis?, *Journal of the ACM (JACM)*, **58** (2011), 1–37.

[5] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Processing Letters*, **14** (2007), 707–710.

[6] J. P. Cunningham and Z. Ghahramani, Linear dimensionality reduction: Survey, insights, and generalizations, *The Journal of Machine Learning Research*, **16** (2015), 2859–2900.

[7] J. F. P. Da Costa, H. Alonso and L. Roque, A weighted principal component analysis and its application to gene expression data, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **8** (2009), 246–252.

[8] F. De la Torre and M. J. Black, Robust principal component analysis for computer vision, in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, IEEE, 2001, 362–369.

[9] E. Elhamifar and R. Vidal, Sparse subspace clustering: Algorithm, theory, and applications, *IEEE transactions on pattern analysis and machine intelligence*, **35** (2013), 2765–2781.

[10] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American statistical Association*, **96** (2001), 1348–1360.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge university press, 2013.

[12] X.-L. Huang, L. Shi and M. Yan, Nonconvex sorted ℓ_1 minimization for sparse approximation, *Journal of the Operations Research Society of China*, **3** (2015), 207–229.

[13] G. Li and T. K. Pong, Global convergence of splitting methods for nonconvex composite optimization, *SIAM Journal on Optimization*, **25** (2015), 2434–2460.

[14] H. Li and Z. Lin, Accelerated proximal gradient methods for nonconvex programming, in *Advances in Neural Information Processing Systems*, 2015, 379–387.

[15] Z. Lin, M. Chen and Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. 2010, arXiv preprint `arXiv:1009.5055`, (2010), 663–670.

[16] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu and Y. Ma, Robust recovery of subspace structures by low-rank representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35** (2012), 171–184.

[17] X. Liu, Z. Wen and Y. Zhang, An efficient Gauss–Newton algorithm for symmetric low-rank product matrix approximations, *SIAM Journal on Optimization*, **25** (2015), 1571–1608.

[18] Y. Lou and M. Yan, Fast l1–l2 minimization via a proximal operator, *Journal of Scientific Computing*, **74** (2018), 767–785.

[19] N. Sha, M. Yan and Y. Lin, Efficient seismic denoising techniques using robust principal component analysis, in *SEG Technical Program Expanded Abstracts 2019, Society of Exploration Geophysicists*, 2019, 2543–2547.

[20] Y. Shen, H. Xu and X. Liu, An alternating minimization method for robust principal component analysis, *Optimization Methods and Software*, **34** (2019), 1251–1276.

[21] M. Tao and X. Yuan, Recovering low-rank and sparse components of matrices from incomplete and noisy observations, *SIAM Journal on Optimization*, **21** (2011), 57–81.

[22] L. N. Trefethen and D. Bau III, *Numerical linear algebra*, vol. 50, SIAM, 1997.

[23] F. Wen, R. Ying, P. Liu and T.-K. Truong, Nonconvex regularized robust PCA using the proximal block coordinate descent algorithm, *IEEE Transactions on Signal Processing*, **67** (2019), 5402–5416.

[24] Z. Wen, W. Yin and Y. Zhang, Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm, *Mathematical Programming Computation*, **4** (2012), 333–361.

[25] J. Wright, A. Ganesh, S. Rao, Y. Peng and Y. Ma, Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization, in *Advances in Neural Information Processing Systems*, 2009, 2080–2088.

[26] X. Yuan and J. Yang, Sparse and low-rank matrix decomposition via alternating direction methods, preprint, 12 (2009).

[27] C.-H. Zhang, Nearly unbiased variable selection under minimax concave penalty, *The Annals of Statistics*, **38** (2010), 894–942.

Received February 2020; 1st revision July 2020; 2nd revision August 2020.

E-mail address: shaningy@msu.edu
E-mail address: leishi@fudan.edu.cn
E-mail address: myan@msu.edu